

Opinion Writing and Authorship on the Supreme Court of Canada

Kelly Bodwin, Jeffrey S. Rosenthal and Albert H. Yoon*

Revised: August 2012

Abstract

In contrast to other branches of government, the Supreme Court of Canada operates with relatively lean staffing. For most of the Court’s history, its justices alone determined which cases to review, heard oral argument, and wrote opinions. Only since 1967 have justices have been aided in these responsibilities by law clerks. While interest abounds in the relationship between justices and their clerks – particularly the writing of opinions – very little is known. This article analyzes the text of the Court’s opinions to better understand judicial authorship. We find that justices possess distinct writing styles, allowing us to distinguish them from one another. Their writing styles also provide insight into how clerks influence the writing of opinions. Most justices in the modern era possess a more variable writing style than their predecessors, both within and across years, providing strong evidence that clerks are increasingly involved in the writing of judicial opinions.

* University of North Carolina, Department of Statistics; University of Toronto, Department of Statistics; University of Toronto Faculty of Law, respectively. Order of authors listed in alphabetical order. We would like to thank Ben Alarie, Ian Caines, Andrew Green, Helen Levy, Ed Morgan, Todd Peppers, Simon Stern, and Michael Tribelcock, and two anonymous reviewers for the *University of Toronto Law Journal*. Rosenthal was supported in part by NSERC, and Yoon was supported in part by the Russell Sage Foundation. All remaining errors are our own.

I. INTRODUCTION

“[A]mbitious judges realize that law clerks help them attain their ambitions. People want the best ghostwriters.”
- Judge Richard Posner¹

The Supreme Court of Canada stands in stark contrast to Parliament and the Prime Minister. The latter includes a large number of elected officials, and legions of staff to help execute their duties. The Court, by comparison, is parsimonious, consisting of only nine justices. With the aid of only their law clerks, the justices are charged with the responsibility of deciding which among hundreds of applications for leave to appeal to hear, and to produce written opinions of every judgment. Each justice is authorized to hire no more than three clerks,² typically in their twenties³ and recent graduates from elite⁴ Canadian law schools

While scholars have long been fascinated with the Court generally,⁵ relatively little, if any, has been written about the subject of authorship for the Supreme Court of Canada, which stands in sharp contrast to the United States Supreme Court, replete with anecdotal accounts of which justices did and did not write their own opinions.⁶ For example, in the

1 David Margolick, “At the Bar: Annual Race for Clerks Becomes a Mad Dash with Judicial Decorum Left in the Dust,” N.Y. TIMES (March 17, 1989), at B4.

2 The information provided by the Supreme Court of Canada web site, available at <http://www.scc-csc.gc.ca/court-cour/administration/index-eng.asp>; *see also* Lorne Sossin, *The Sounds of Silence: Law Clerks, Policy Making and the Supreme Court of Canada*, 30 U. BRIT. COLUM. L. REV. 279, 283 (1996) (describing the number of clerks in each chamber);

3 *See* F.L. MORTON, RAINER KNOPFF, *THE CHARTER REVOLUTION AND THE COURT PARTY* 110 (2000) (describing the burgeoning power of law clerks at the Supreme Court of Canada).

4 For example, McInnes et al. report that over 75% of clerks during the early 1990s attended one of seven law schools (Toronto, McGill, Osgoode Hall, Ottawa, Dalhousie, Alberta, and the University of British Columbia). *See* Mitchell McInnes, Janet Bolton & Natalie Derzko, *Clerking at the Supreme Court of Canada*, 33 Alberta L. R. 58 (1994).

5 *See e.g.*, DONALD R. SONGER, *THE TRANSFORMATION OF THE SUPREME COURT OF CANADA: AN EMPIRICAL EXAMINATION* (2008).

6 *See, e.g.*, Transcriptions of Conversations between Justice William O. Douglas and

United States, Justices Holmes, Cardozo, and Douglas, by reputation, wrote their own opinions.⁷ Latter justices, such as Justices Marshall, by comparison, were reputed to have relied more on their clerks.⁸ These anecdotal accounts of authorship are strongly supported by recent statistical analyses of their own writing.⁹

Relatively little is known about the role of judicial clerks in the Supreme Court of Canada. Some scholars assert that judicial clerks in Canada have grown in power, rivaling those at the United States Supreme Court.¹⁰ Clerks not only assist in evaluating appeals and writing bench memos, they are alleged to have created new judicial doctrines on behalf of their justices.¹¹ Others contend that the justices do their own writing.¹² Lorne Sossin, Dean of Osgoode Hall School of Law and a former Supreme Court clerk, contends that “clerks now play a more significant role in the institution of the Supreme Court than ever before,” but maintain that “[c]lerks do what they are told to do and have

Professor Walter F. Murphy, Cassette No. 3: December 20, 1961, Princeton University Library, 1981 (where Justice Douglas stated “I have written all my own opinions.”); EDWARD LAZARUS, *CLOSED CHAMBERS* 271 (1998) (identifying Justices Scalia and Stevens as writing their own opinions).

⁷ See TODD C. PEPPERS, *COURTIERS OF THE MARBLE PALACE: THE RISE AND INFLUENCE OF THE SUPREME COURT LAW CLERK* 58, 96-97, 114 (2006) (describing the relationship of Justices Holmes, Cardozo, and Douglas to their clerks).

⁸ See Mark Tushnet, *Thurgood Marshall and the Brethren*, 80 *GEO. L. J.* 2109, 2112 (1992) (describing how Marshall relied more heavily on his law clerks early in his tenure).

⁹ See Jeffery S. Rosenthal & Albert H. Yoon, *Judicial Ghostwriting: Authorship on the U.S. Supreme Court*, 96 *Cornell L. Rev.* 1307, 1323-24 (showing in Figure 2 the year-to-year variability scores of these, among other, justices).

¹⁰ See MORTON & KNOPFF, *supra* note 3, at 110 (describing the burgeoning power of law clerks at the Supreme Court of Canada).

¹¹ See *id.* at 111 (describing how a law clerk for Chief Justice Dickson purportedly created Section 1 of the Oakes Test).

¹² See McInnes, *supra* note 4, at 78 (writing that “any suggestion that Canada’s Justices abdicate their responsibilities for writing judgments is false; in every instance, a decision of the Court ultimately and fundamentally is the product of the nine men and women who sit on the bench.”)

no independent duties or tasks beyond the instructions they receive from the Justices.”¹³ Other scholars echo this viewpoint, having written that justices “have not surrendered considerable discretion and initiative to their own clerks.”¹⁴ While recent scholarship provides evidence that older workers may be more productive than their junior coworkers,¹⁵ the Court is unique in that it places a heavy workload on a relatively few number of older individuals.

Given the prominence of the Court, the importance of judicial authorship extends beyond mere academic inquiry. First, the Court provides the final say on judicial matters in Canada. Second, in a common law regime where judicial precedent matters, the reasoning that the Court applies in an opinion is arguably just as important as the prevailing party. Lower courts, practitioners, and legal scholars carefully deconstruct the reasoning of the Court. For the reason, the extent to which the reasoning of the Court reflects the justices thinking – or those of their clerks - matter.

At the same time, the institutional design of the Court provides a natural identification strategy to test our hypothesis. Supreme Court clerkships are typically for a single term, running from October through August of the following year. If justices indeed depend heavily on their clerks, the annual turnover in judicial clerks should reveal this reliance. Our premise: justices who rely more on their clerks in the opinion-writing process possess a more variable writing style than their less reliant colleagues, both within and across years. Moreover, the relatively recent inception of judicial clerks on the Court in 1967 suggests that the writing variability in the pre-clerk years should be

13 See Sossin, *supra* note 2, at 297-98.

14 See PETER MCCORMICK & IAN GREENE, *JUDGES AND JUDGING: INSIDE THE CANADIAN JUDICIAL SYSTEM* 208 (1990). This viewpoint is echoed by Justice Bertha Wilson, *see* Bertha Wilson, *Decision-Making in the Supreme Court*, 36 U. Tor. L. J. 227, 236 (1986) (describing how clerks central role is to provide background research).

15 See, e.g., Axel Börsch-Supan & Matthias Weiss, *Evidence from Work Teams at the Assembly Line*, working paper (2011), available at http://www.mea.uni-mannheim.de/uploads/user_mea_discussionpapers/1057_MEA-DP_148-2007.pdf (finding that productivity among auto workers does not decline as they grow older, suggesting that any physical decline may be offset by experience and an ability to collaborate with others); Ray C. Fair, *How Fast Do Old Men Slow Down?*, 76 Rev. Econ. & Stat. 103-18 (1994) (describing physical decline among professional athletes).

lower than the post-clerk years. The progression towards multiple clerks within each chamber should, for those justices relying heavily on their clerks in the opinion-writing process, further increase their writing variability.

In this article, we analyze the text of majority opinions written by all justices on the Court. We construct a variability score based on the justices' own use of common function words (e.g., *for*, *have*, *with*). We find that each justice possesses a unique variability score. More importantly, however, even justices with nearly identical variability scores possess distinct writing styles based on these function words. These differences in turn allow us to correctly identify authorship in a pairwise comparison of justices. We find that current and recent justices possess higher variability scores than the justices preceding them, providing statistical evidence that justices, on average, rely more on their clerks when writing opinions.

This article proceeds as follows. We describe in Part II our methodology of measuring writing variability through the use of common function words (e.g., *an*, *have*, *what*). In Part III we describe our data of the Court's decisions. We report our main results in Part IV, showing the justices' variability scores, both within and across cohorts, and our ability to use these scores to accurately predict authorship in pairwise comparisons of justices. We discuss the implications of our results in Part V.

II. METHODOLOGY

In analyzing the text of judicial opinions from the Supreme Court of Canada, we follow the methodology established by Rosenthal and Yoon in their analysis of the United States Supreme Court.¹⁶ We briefly summarize the approach here.

Judicial writing, and writing in general, is influenced on two factors. The first is the subject or topic of the writing; the second is the author's writing style, i.e., her word choice (diction) and sentence structure (syntax). Our statistical analysis focuses on the latter factor, falling within the broader discipline of stylometry.¹⁷ Specifically, we

¹⁶ See Rosenthal & Yoon, *supra* note 9, at 1313-1317 (2011).

¹⁷ Scholars have applied stylometry to literature, *see e.g.*, O. Seletsky et al, *The Shakespeare Authorship Question*, unpublished manuscript, Dartmouth College (2007);

analyze commonly used words, often referred to as function words.¹⁸ Our central assumption is that an author's use of such words operates independently of the subject or topic of the writing.¹⁹

The intuition behind our approach is that individuals' writing styles tend to be fairly consistent; thus, greater variability in writing style reflects that a justice delegates more of her writing responsibilities to her law clerks: the more people actively involved in writing judicial opinions, all things equal, the more variable the writing style. For example, if Justice A uses the word “this” about 2% of the time and the word “some” about 3% of the time, then opinions authored solely by Justice A will tend to follow these percentages fairly consistently and thus have fairly low variability scores. Similarly, if Justice B uses the word “this” about 1% of the time and the word “some” about 4% of the time, then opinions authored solely by Justice B will follow these percentages fairly consistently and again have fairly low variability scores. In particular, solely-authored items are characterized by low *variability* of word frequencies, not by the specific word frequencies themselves.

On the other hand, if Justice A instead lets many different law clerks write her different opinions, then some of those law clerks might use “this” much more than 2% of the time, and others might use “this” much less than 2% of the time, and others might use “some” much more than 3% of the time, and so on. If so, then Justice A's opinions' word frequencies will change significantly from one opinion to the next, and will thus have much higher variability scores. (Of course, it is conceivable that some of the law clerks will happen to have very similar word usage to Justice A herself, but it is highly unlikely that many of the clerks will be very similar in all 63 function words that we consider.) In

political texts, *see e.g.*, FREDERICK MOSTELLER & DAVID L. WALLACE, *INFERENCE AND DISPUTED AUTHORSHIP: THE FEDERALIST* (1964); and speeches, *see e.g.*, E.M. Airoidi, et al. *Who Wrote Ronald Reagan's Radio Addresses?* 1 *BAYESIAN ANALYSIS* 289 (2006).

18 Other approaches of writing style – e.g., sentence length, paragraph length, punctuation – yielded comparable results.

19 D. Madigan, et al, *Author Identification on the Large Scale*, Proceedings of the Classification Society of North America (CSNA) (2005) (manuscript on file with authors).

this way, high variability scores can potentially detect cases where different law clerks are writing different opinions.

We use Mosteller & Wallace's²⁰ original list of 70 function words, and delete seven words rarely used by the justices.²¹ The remaining 63 words appear in Table 1.

TABLE 1
STYLOMETRY OF SUPREME COURT TEXT
63 FUNCTION WORDS

(1-12) *a, all, also, an, and, any, are, as, at, be, been, but,*
(13-24) *by, can, do, down, even, for, from, had, has, have, her, his,*
(25-36) *if, in, into, is, it, its, may, more, must, no, not, now,*
(37-48) *of, on, one, only, or, our, so, some, such, than, that, the,*
(49-60) *their, then, there, things, this, to, up, was, were, what, when,*
which,
(61-63) *who, with, would*

For each opinion, we keep count of each of the aforementioned words. Given this approach, we adopt a chi-squared approach to test the distribution of the word counts we observe against a null hypothesis that, for each justice, the total count of each function word is equally likely to occur in any of the total number of her opinions.

The chi-squared statistic is the following:

$$chisq = \sum_{i=1}^K \sum_{j=0}^{63} \frac{(c_{ij} - e_{ij})^2}{e_{ij}}$$

where j are the function words numbered from $j=1$ to $j=63$; K represents the total number of opinions a justice has written in our dataset, numbered from $i=1$ to $i=K$; c_{ij} represents the number of times that function word j appears in opinion i . The term e_{ij} is the expected number of times that function word j would have appeared in judgment i , and w_i be the total number of words in judgment i .

We calculate our variability score by the following:

²⁰ See MOSTELLER & WALLACE, *supra* note 17.

²¹ The seven words – *every, my, shall, shold, upon, will, you* – each constituted fewer than 0.1% of all words in the Court's majority opinions.

$$\text{Variability Score} = \frac{\text{chisq}}{\text{df}} = \frac{\text{chisq}}{63(K-1)}$$

where df is the degrees of freedom. This variability score provides the foundation of all subsequent analysis.

To test the null hypothesis that the justices' writing style follows a uniform and random distribution of function words, we randomly generate 200 pseudo-documents, each consisting of 2000 independently and randomly generated words. We set up the documents such that each of the 2000 words had a 70% probability of being a nonfunction word, and 30% probability of being a function word from Table 1. The score approximates 1 under the null hypothesis. We repeated this experiment 10 times, which produces a mean variability score of 1.004622 with a standard deviation of 0.001702, consistent with a null hypothesis with a true mean equal to one. A variability score that exceeds 1 reflects greater variability. In other words, the higher the variability score, the more variable her writing style, based on the function words in Table 1.

Finally, because chi-squared values are less stable when expected cell counts for given words approximate zero, we exclude function words with a very low frequency, as well as opinions shorter than 250 words.

Our central identification strategy for discerning writing variability is law clerks themselves. Law clerks typically serve for only one year, replaced annually by a new set of law clerks. If law clerks vary in their own writing style, then justices who rely more on their clerks for drafting or writing of opinions would reveal greater variability in writing style than justices who do their own writing.

While we believe that this identification strategy provides a compelling proxy for how much justices rely on their law clerks, we recognize that it is ultimately untestable. An alternative explanation always exists that even in a world where all justices wrote their own opinions, some justices have more variable writing styles than others. The relationship between justices and their clerks remain closely guarded, in the United States²² and even more so in Canada. But we maintain that greater variability in writing

²² See, e.g., PEPPERS, *supra* note 7, at 18-20 (describing clerks' reluctance to discuss their relationship with their justices, citing confidentiality).

both within and across years strongly suggest greater reliance by justices on their law clerks. As an additional validity check, two prominent United States Courts of Appeal judges – Richard A. Posner and Frank Easterbrook – widely reputed to write their own opinions, exhibit variability measures that are both markedly lower and less variable than contemporary United States justices.²³

III. DATA

We constructed a dataset of the written judgments from the Supreme Court of Canada, available at scc.lexum.org. This website – a collaborative effort of the Court, Lexum, and the University of Montreal – includes all published opinions since the inception of the Court, dating back to 1876. While the dataset spans the entire history of the Court – comprising 76 justices²⁴ – our analysis focuses on justices in the period after 1900.

We analyzed the opinions using software in C and Unix that downloaded the decisions directly from the Lexum website. The opinions for each case were contained in HyperText Markup Language (HTML) format.²⁵ Specifically, the program converts the HTML pages into plain text and extracts from each opinion text not written by the justices themselves, including headnotes and procedural synopses (though quotations within the text are not removed). Our program also separates majority from concurring and dissenting opinions. The resulting text files were then spot checked to ensure accuracy and avoid downloading errors.

23 Judge Posner’s variability measure was 2.60 for the period 1981-2010 with a standard deviation of 0.14, while Judge Easterbrook’s was 2.42 with a standard deviation of 0.18. *See* Rosenthal & Yoon, *supra* note 16, at 1325. Their variability scores were lower than any of their contemporary justices (for those on the Court as of 2010) on the U.S. Supreme Court, which range from 3.06 (J. Breyer) to 3.73 (J. Kennedy); their standard deviations were similarly lower than those of the justices, which ranged from 0.56 (J. Stevens) to 0.62 (J. Kennedy). *See id.* at 1323.

24 For justices who joined the Court prior to 1949, see Appendix, Table A.

25 For the software used to download and analyze the text, and a description of the software program, see Jeffrey S. Rosenthal, *Explanation of the Software, Probability.ca*, <http://probability.ca/scc/README>.

Although we analyze each type of opinion within each judgment – majority, concurring, dissenting – we focus on majority opinions. Concurring and dissenting opinions have become more common, but most judgments today consist just of a single, majority opinion. Because dissents and concurrences are often shorter than majority opinions, methodologically they create instability in the textual analysis. We also exclude *per curiam* opinions and co-authored opinions, both sparingly used by the Court.²⁶

IV. RESULTS

Our analysis in the results builds from our variability score. We first establish that the justices' variability scores are statistically distinguishable from the null hypothesis. We then construct a bootstrap approach to determine whether, in a pairwise comparison, justices' variability scores are statistically distinguishable from one another. To test whether our variability score can accurately predict authorship, we design a linear classifier, enabling us to compare scores for each justice, looking in particular at the beginning and end of their tenure. Third, we test the accuracy of the linear classifier through a *leave-one-out cross-validation*. As mentioned earlier, our analysis, unless otherwise stated, is based on majority opinions.

Variability Scores: We begin by describing the variability scores the justices on the Supreme Court of Canada. These scores provide a measure of the each justice's writing style, from which we can compare to one another. Table 2 lists the 38 justices who joined the court after 1948, including the province they represent, the prime minister who appointed them, when they joined and left the Court, how many majority opinions they wrote, the average length of their majority opinions, and importantly, their variability score.

²⁶ See Charrissima Mathen, *The Coming of Age of the Charter Dissent and Judicial Authority in Charter Cases*, 15 U.N.B. L.J. 321, 323 (noting that the Court “rarely issues *per curiam* decisions”). Our analysis of the data reveals that co-authored opinions similarly occur infrequently.

A higher than average variability score, as noted earlier, is consistent with the view that the given justice has relied more on her law clerks in writing opinions. It bears repeating that we acknowledge the possibility of an alternative explanation that even in a world where justices did their own writing, some justices possess a greater variability in writing style than others. It is also possible that even in a world where all justices relied on clerks to write opinions, some clerks are better than others at mimicking their justices' writing styles. Thus, the variability measure could be capturing differences in clerks' abilities. Because of the incompatibility of the textual analysis, in this table and throughout our analysis we exclude justices who wrote their opinions in French.²⁷

²⁷ For a comparison of variability scores by the justices' original language of opinion, see Figure A2 in the appendix.

TABLE 2
VARIABILITY SCORE
CURRENT AND DEPARTED JUSTICES (JOINED POST-1948)

Num	Justice	Province	Appointing Prime Minister	Year Joined Court	Year Left Court	Years On Court	Opinions	Average Word Length	V4 Score
1	Thomas Cromwell	NS	Harper	2008	present	3	14	5748	4.75
2	Marshall Rothstein	MB	Harper	2006	present	5	39	6551	4.60
3	Louise Charron	ON	Martin	2004	present	7	45	6312	4.10
4	Rosalie Abella	ON	Martin	2004	present	7	52	4844	4.27
5	Morris J. Fish	QC	Chrétien	2003	present	8	55	3758	3.00
6	Marie Deschamps	QC	Chrétien	2002	present	9	48	7371	4.25
7	Louis LeBel	QC	Chrétien	2000	present	11	115	7367	4.43
8	Louise Arbour	ON	Chrétien	1999	2004	5	47	7194	4.13
9	William Ian Corneil Binnie	ON	Chrétien	1998	present	13	109	7733	3.79
10	Michel Bastarache	NB	Chrétien	1997	2008	11	81	7748	4.41
11	John C. Major	AB	Mulroney	1992	2005	13	99	5161	4.33
12	Frank Iacobucci	ON	Mulroney	1991	2004	13	123	8783	5.02
13	William Stevenson	AB	Mulroney	1990	1992	2	14	3108	2.84
14	Beverley McLachlin	BC	Mulroney	1989	present	22	290	5439	4.56
15	Peter Cory	ON	Mulroney	1989	1999	10	121	6173	4.20
16	Charles Gonthier	QC	Mulroney	1989	2003	14	102	7459	4.94
17	John Sopinka	SK	Mulroney	1988	1997	9	145	4476	3.17
18	Claire L'Heureux-Dubé	QC	Mulroney	1987	2002	15	68	6496	4.24
19	Gérard La Forest	NB	Mulroney	1985	1997	12	141	6721	3.77
20	Gerald Eric Le Dain	ON	Trudeau	1984	1988	4	31	5531	3.15
21	Bertha Wilson	ON	Trudeau	1982	1991	9	112	5736	3.97
22	Antonio Lamer	QC	Trudeau	1980	2000	20	228	5542	4.32
23	Julien Chouinard	QC	Clark	1979	1987	8	56	4234	3.14
24	William Rogers McIntyre	BC	Trudeau	1979	1989	10	108	4068	3.16
25	Yves Pratte	QC	Trudeau	1977	1979	2	24	4436	3.10
26	Louis-Philippe de Grandpré	QC	Trudeau	1974	1977	3	71	2435	2.62
27	Jean Beetz	QC	Trudeau	1974	1988	14	82	6807	4.47
28	Robert George Brian Dickson	MB	Trudeau	1973	1990	17	213	5432	3.72
29	Bora Laskin	ON	Trudeau	1970	1984	14	245	3282	3.04
30	Louis-Philippe Pigeon	QC	Pearson	1967	1980	13	206	2727	2.78
31	Wishart Flett Spence	ON	Pearson	1963	1978	15	166	3272	3.19
32	Emmett Matthew Hall	SK	Diefenbaker	1962	1973	11	94	2456	2.98
33	Roland Almon Ritchie	NS	Diefenbaker	1959	1984	25	286	2856	2.81
34	Wilfred Judson	ON	Diefenbaker	1958	1977	19	211	1708	2.62
35	Ronald Martland	AB	Diefenbaker	1958	1982	24	274	2999	3.81
36	Henry Grattan Nolan	AB	St. Laurent	1956	1957	1	9	2950	2.69
37	Douglas Charles Abbott	QC	St. Laurent	1954	1973	19	121	1183	2.73
38	John Robert Cartwright	ON	St. Laurent	1949	1970	21	271	2280	2.82

Note: Justices are arranged in reverse chronological order, by year of appointment. Justices who were appointed to the court prior to 1949 are listed in the Appendix, Table A1. This table excludes justices who wrote their opinions in French.

The variability scores range from a high of 5.02 for Justice Frank Iacobucci to a low of 2.62 for Justice Louis-Philippe de Grandpré. The variability score for each justice is statistically significantly larger than the null hypothesis of uniformly and randomly

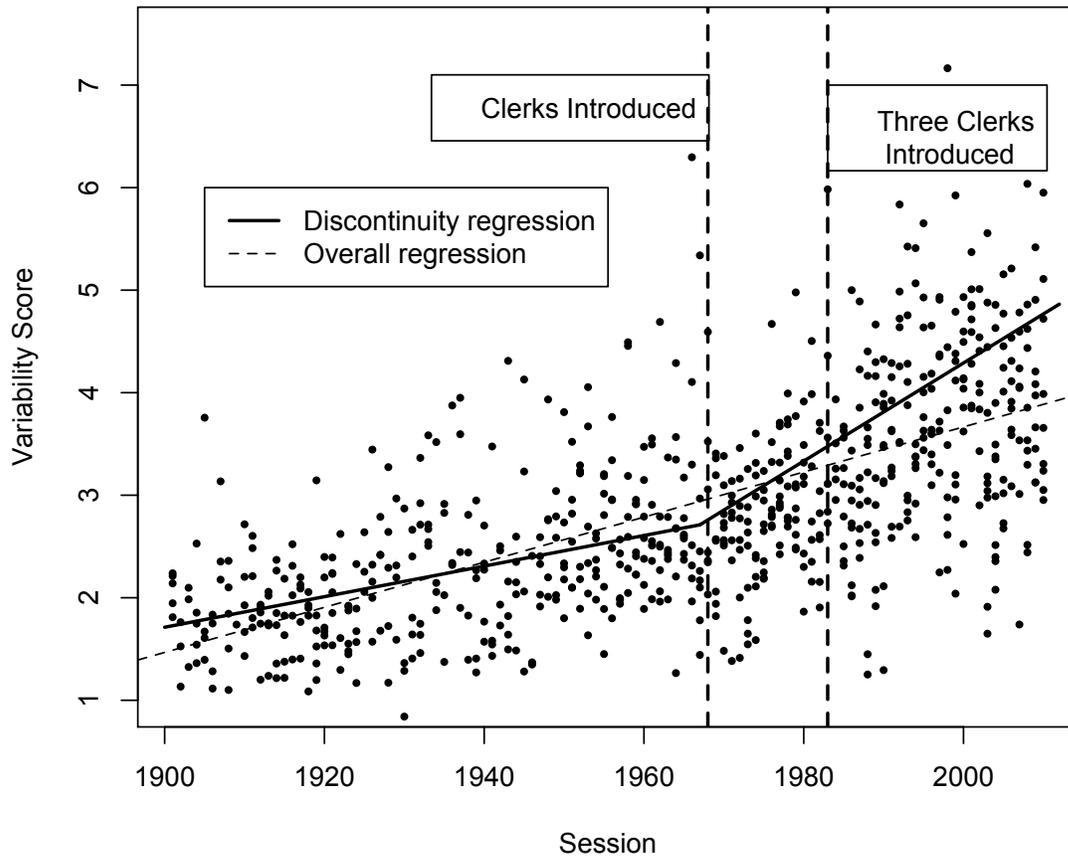
distributed function words. As an example, Justice Louis LeBel’s variability score of 4.43 reflects a *chisq* statistic of $4.43 \times 63 \times (114-1) = 31,537.17$. The null hypothesis has a chi-squared statistic of $63 * (114-1) = 7,119$. The *p* value corresponding to Justice LeBel’s score is less than 0.00001, allowing us to reject the null hypothesis. For close followers – and perhaps current and former law clerks – of the Court, the variability scores may support or contradict their beliefs about the extent to which justices write their own opinions. Collectively, Table 2 suggests that the variability in writing has increased over time across justices.

Figure 1, showing variability scores for justices dating beginning in 1900, graphically shows this upward trend. The sloped dotted line represents the regression line over the entire period, with a significant positive slope ($p < 0.001$). Although justices sitting together differ in their writing variability, the regression line is positively sloped over time.

It is worth noting the differences in slope before and after the inception of law clerks in 1967, as represented by the thick black line in Figure 1. For the period 1900-1966, the regression line is flatter relative to the regression line for the entire period, with the Variability Score increasing by about 0.15 per year. Conversely, for the period 1967-2010, the regression line is steeper relative to the overall regression line, increasing at a rate of 0.33 per year.²⁸ Interestingly, attempts to fit a jump discontinuity at 1967 (as opposed to the kinked regression in Figure 1) did not lead to a statistically significantly better fit. This means there was no immediate change in V4 scores with the advent of clerks; rather, one might infer that justices gradually adapted to the presence of clerks and their Variability Scores increased more quickly as a result. It is also of note that no significant jump or change in slope was found at 1983, the year of expansion to a three clerk system. Perhaps, then, the inflation of clerks was not an exogenous shock, but rather a response to the justices' evolving reliance upon their assistants.

²⁸ Both quoted rates are significantly distinct from 0, and from each other, at $p < 0.001$..

FIGURE 1
VARIABILITY SCORES OVER TIME



Note: Sloped dotted line represents regression line over entire period. Sloped solid lines represent regression line within each time period.

Similarly, Table 3 shows an upward trend in writing variability when categorizing time periods by the presiding Chief Justice. This evidence illustrates that justices' writing style was less variable prior to law clerks, and became increasingly more variable following the introduction of law clerks.

TABLE 3
VARIABILITY SCORE
BY CHIEF JUSTICE

Justice	Province	Elevating Prime Minister	First Year as Chief	Last Year as Chief	Total Years on Court	Total Opinions	Average Word Length	Variability Score
Beverley McLachlin	BC	Chrétien	2000	present	22	807	6788	4.61
Antonio Lamer	QC	Mulroney	1990	2000	20	1194	6190	4.36
Robert George Brian Dickson	MB	Trudeau	1984	1990	17	541	5398	3.79
Bora Laskin	ON	Trudeau	1973	1984	14	1179	3563	3.43
Gérard Fauteux	QC	Trudeau	1970	1973	24	392	2502	3.06
John Robert Cartwright	ON	Pearson	1967	1970	21	354	2290	3.18
Robert Taschereau	QC	Pearson	1963	1967	27	383	2195	3.19
Patrick Kerwin	ON	St. Laurent	1954	1963	28	854	2100	3.04
Thibaudeau Rinfret	QC	King	1944	1954	30	605	2182	2.96
Sir Lyman Poore Duff	BC	Bennett	1933	1944	38	232	2256	3.01
Francis Alexander Anglin	ON	King	1924	1933	24	192	2063	2.76
Sir Louis Henry Davies	PEI	Borden	1918	1924	23	216	1497	2.39
Sir Charles Fitzpatrick	QC	Laurier	1906	1918	12	305	1682	2.28
Sir Henri Elzéar Taschereau	QC	Laurier	1902	1906	28	105	1536	2.43
Sir Samuel Henry Strong	ON	Thompson	1892	1902	27	204	2028	3.01
Sir William Johnstone Ritchie	NB	Macdonald	1879	1892	17	255	2557	3.35
Sir William Buell Richards	ON	NA*	1875	1879	4	49	3200	4.08

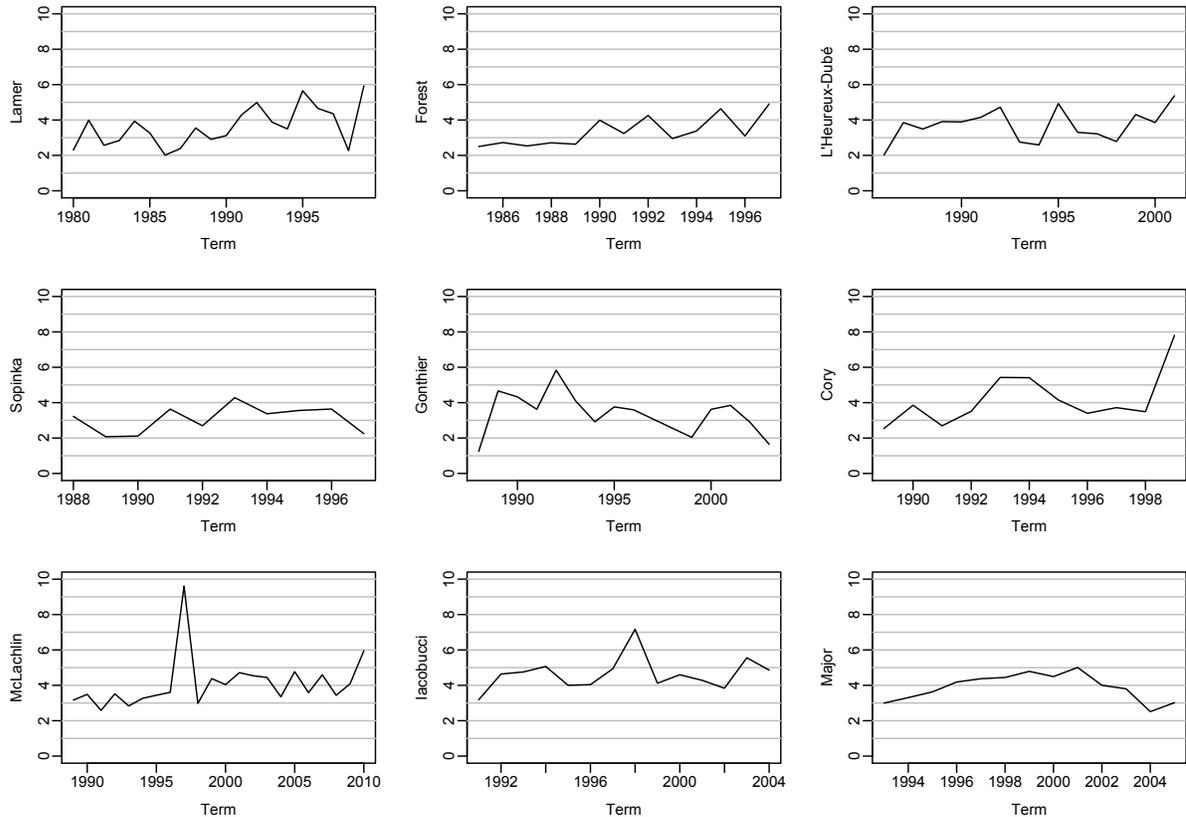
Note: * Variability (V) analysis not performed for francophone justices. ** Appointed as Chief Justice upon creation of the Court. Table excludes justices who wrote their opinions in French.

These aggregate scores, while informative, are incomplete. An aggregate score, by itself, does not reveal how the justices' scores may vary over their tenure on the Court. Theoretically, there are reasons to suggest both how they may increase and decrease. The job is intellectually demanding and may impose a steep learning curve at the beginning. Also, as justices become older, their willingness to delegate writing obligations may increase.

Figure 2 provides a sample of recent justices throughout their tenure. Justice Forest, for example, with an aggregate variability score of 3.77, had lower – and remarkably consistent – writing variability during his first five years on the Court. His writing style subsequently became more variable, both within and across years. Justices Lamer, L'Heureux-Dubé, Sopinka, Cory, McLachlin, and Iacobucci similarly follow an increasing variability in writing style during their tenures. Justice Gonthier, by contrast, was one of the few justices whose writing style became less variable over the bulk of his tenure, while Justice Major's writing style steadily grew more variable during the first

eight years on the bench but became less variability during his final five years on the Court.

FIGURE 2
YEAR-TO-YEAR VARIABILITY SCORES
SELECT JUSTICES

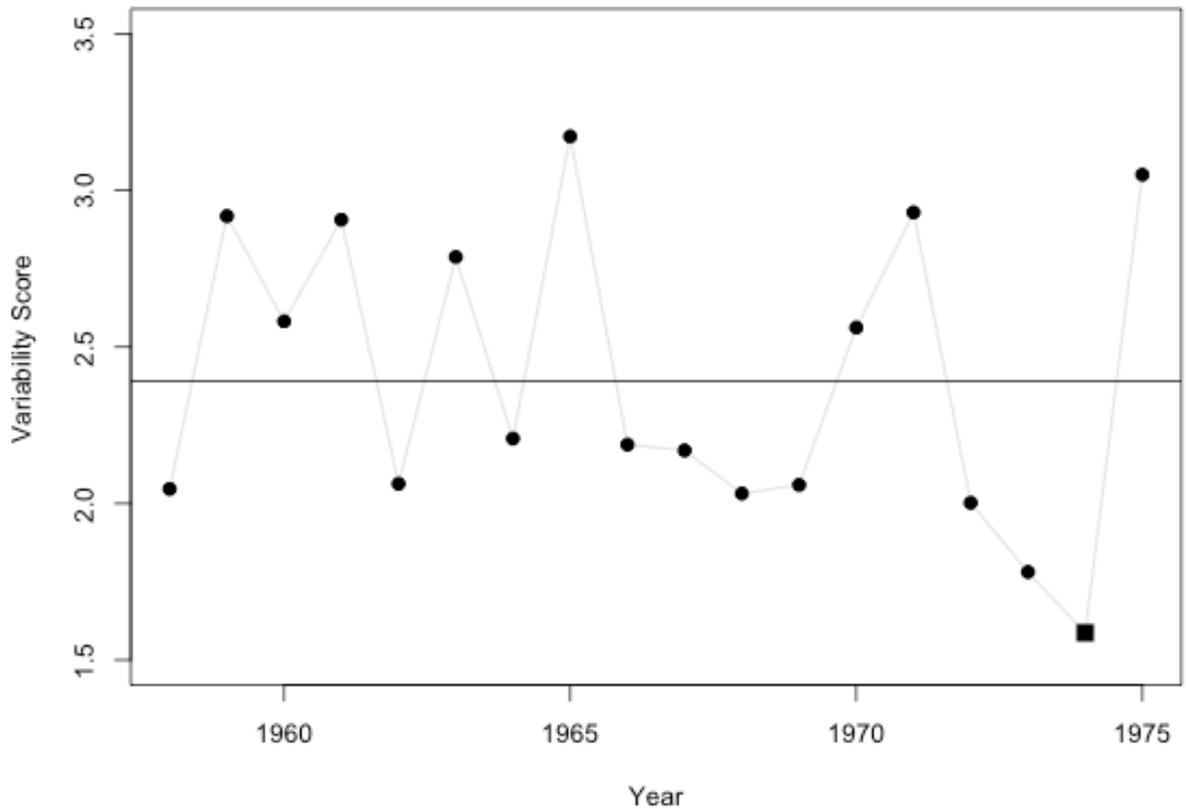


As stated in the methodology section, law clerks serve as the central identification strategy for discerning writing variability. One limitation of this approach is that post-1967, all justices are reputed to have hired law clerks, meaning that there is no reason to expect variation across justices subsequent to 1967. Justice Judson provides an exception to the institutional trend. For the 1974-75 term, Justice Judson did not hire a law clerk,²⁹ meaning that he wrote his own opinions. Figure 3 supports the view that the presence and absence of a law clerk affects writing variability. Following the implementation of

²⁹ Justice Judson did not have a law clerk during the 1974-75 term. See Michael J. Herman, *Law Clerking at the Supreme Court of Canada*, 13 OSGOODE HALL L. J. 279 (1975) (reporting this fact).

law clerks, Justice Judson’s writing variability increased in the short term (1969-70 through 1971-72), before declining. While one should exercise caution against over-interpretation of these results, given the variation over time, Justice Judgson’s variability score in the 1974-75 term was the lowest of his career. Moreover, for the 1975-76 term, after Justice Judson hired a law clerk, his writing variability sharply increased.

FIGURE 3
VARIABILITY SCORE
JUSTICE JUDSON



Note: Square marker indicate year in which Justice Judson did not have a law clerk.

Significance Tests: Our variability measure for each of the justices allow us to reject the null hypothesis that each justice’s writing style follow a uniform and random distribution of function words. It does not, however, speak to whether the differences in variability scores from one justice to another are statistically meaningful. For this question, no simple analytic test exists. Because we rejected the null in each instance, the

justices' scores by definition do not follow a chi-squared distribution. To complicate matters, it is not possible to analytically determine the shape of the distribution.

To get around this analytical intractability, we use a bootstrap test: repeated sampling with replacement from a given sample,³⁰ which enables us to empirically determine the distribution. We randomly select for each justice 100 authored majority opinions, with repetition.³¹ We compute the variability score for this sample. We then repeat the process 1000 times for each justice, generating 1000 different possible variability scores, depending on the opinions selected for each sample.

We then do pairwise comparisons of justices, based on their bootstrap variability scores. This process creates 1 million (1000 x 1000) pairs of variability scores. We then tally the number of pairs where the variability score is greater for Justice A than for Justice B. The resulting fraction provides an estimate of the probability that Justice A's variability score is greater than Justice B's variability score for a random selection of judgments. From these pairings we can estimate the distribution function for the difference in variability scores between the two justices, from which we can compute a 95% confidence interval. As with any confidence interval, a wholly positive or negative confidence interval indicates that the difference in variability scores between Justices A and B are statistically meaningful.

Table 4 illustrates two examples of bootstrap pairwise comparisons. Each comparison reports the bootstrap scores for each of the two justices. The bootstrap variability score is similar to the observed variability scores for each justice in Table 2. The greatest divergence is for Justice L'Heureux-Dubé, in large part because she wrote only 68 opinions, far fewer than the other three justices in Table 4. Unsurprisingly, sampling 100 opinions from a pool of 68 opinions will generate greater divergence from the observed variability score. For the first comparison, Justice McLachlin had a higher

30 See BRADLEY EFRON & ROBERT J. TIBSHIRANI, AN INTRODUCTION TO THE BOOTSTRAP (1993) (describing the bootstrapping approach).

31 We include replacement in accordance with convention – see, e.g., BRADLEY EFRON & ROBERT TIBSHIRANI, AN INTRODUCTION TO THE BOOTSTRAP (1993); in addition, because of the relatively low number of opinions written by some justices; sampling without replacement would cause some justices to drop from our analysis.

bootstrap variability than Justice L’Heureux-Dubé. A probability less than 0.05 or greater than 0.95 indicates a statistically significant difference in the bootstrap variability score. The probability that Justice McLachlin’s variability was less than Justice L’Heureux-Dubé was 0.09, which is not statistically significant (at the $p < 0.05$ level). Conversely, Justice Iacobucci’s significantly higher bootstrap score than Justice Sopinka’s bootstrap score was statistically significant.

TABLE 4
VARIABILITY SCORE PAIRWISE BOOTSTRAP
SAMPLE COMPARISONS

Bootstrap Variability Score (McLachlin)	Bootstrap Variability Score (L’Heureux-Dubé)	Probability (McLachlin < L’Heureux-Dubé)	95% C.I. (McLachlin – L’Heureux-Dubé)
4.50	3.91	0.0868	(-1.6236, 0.2139)
Bootstrap Variability Score (Iacobucci)	Bootstrap Variability Score (Sopinka)	Probability (Iacobucci < Sopinka)	95% C.I. (Iacobucci – Sopinka)
4.94	3.12	0.0000	(-2.4715, -1.1360)

Table 5 reports the pairwise comparison of all justices from the 1992-1997 “Lamer natural court” period (i.e., a period when Lamer was chief justice and no other justices joined or left the court). The original variability score for each justice is on the left-most column.³² Each cell reports the probability that the justice listed in the row header has a lower variability score than the justice listed at the column header. For example, the probability that Justice McLachlin has a lower variability score than Justice La Forest is 0.0282, a statistically significant difference. Across the shaded diagonal, the variability scores are mirror images, meaning that the probability that Justice La Forest has a lower variability score than Justice La Forest is 0.9718.

As a general matter, the greater the difference in variability scores, the more likely the bootstrap produces a statistically distinguishable difference. For example, because Justice Iacobucci had the highest variability score (5.02) of this natural court, the

³² Each iteration of the bootstrap score produces a slightly different variability score, but each similar to the original variability score. For this reason, we elect to report the original variability score.

bootstrap comparison of his score was statistically distinguishable from all of the justices, save Justice McLachlin and Justice Gonthier, the two justices with the next highest variability scores. Conversely, Justice Sopinka, by a considerable margin, had the lowest variability score (3.17) in this cohort, statistically distinguishing him from all other justices. In instances where a pair of justices has nearly identical variability scores – as is the case with Justice Lamer (4.32) and Justice Major (4.33) – the bootstrap probabilities converge to 0.50. Accordingly, justices with variability scores near the median of the court were, perhaps unsurprisingly, less distinguishable than justices with relatively low or high variability scores.

Of the 36 unique justice pairings, 18, or 50%, were statistically significant. Even when falling short of statistical significance at the $p < 0.05$ level, 30, or 83%, of these pairings report a pairwise probability less than 0.30 or greater than 0.70.

TABLE 5
VARIABILITY SCORE PAIRWISE BOOTSTRAP
SITTING JUSTICES DURING LAMER NATURAL COURT (1992-97)

	Variability Score	Antonio Lamer	G�rard La Forest	Claire L'Heureux-Dub�	John Sopinka	Charles Gonthier	Peter Cory	Beverley McLachlin	Frank Iacobucci	John C. Major
Antonio Lamer	4.32		0.0764	0.3032	0.0007	0.9269	0.3727	0.6835	0.9457	0.4804
G�rard La Forest	3.77	0.9236		0.8255	0.0270	0.9998	0.8871	0.9718	0.9998	0.9566
Claire L'Heureux-Dub�	4.07	0.6968	0.1745		0.0009	0.9916	0.6124	0.8648	0.9956	0.7522
John Sopinka	3.17	0.9993	0.9730	0.9991		1.0000	0.9997	1.0000	1.0000	1.0000
Charles Gonthier	4.94	0.0731	0.0002	0.0084	0.0000		0.0188	0.2316	0.6013	0.0305
Peter Cory	4.20	0.6273	0.1129	0.3876	0.0003	0.9812		0.7738	0.9865	0.6289
Beverley McLachlin	4.56	0.3165	0.0282	0.1352	0.0000	0.7684	0.2262		0.8026	0.2852
Frank Iacobucci	5.02	0.0543	0.0002	0.0044	0.0000	0.3987	0.0135	0.1974		0.0156
John C. Major	4.33	0.5196	0.2478	0.2478	0.0000	0.9695	0.3711	0.7148	0.9844	

Note: V Score column represents V Score for each Justice. Remaining cells represent the probability that (V Score for Row Justice) < (V Score for Column Justice)

As a separate inquiry, we construct variability scores *within* justices, in an effort to examine how individual justices may have changed over their tenure on the Court,³³ reported in Table 6. Political scientists studying the U.S. Supreme Court have commented how justices shift ideologically while on the Court. The variability measure, because it is based on common function words, should be uncorrelated with ideology. There may, nevertheless, be a connection between ideology and writing variability. To the extent that justices rely more on law clerks in writing opinions, the law clerks' may influence both the substance (ideology) and writing style.

Even for those with prior judicial experience, being a Supreme Court justice is a unique experience that requires a period of adjustment. The Court's caseload, oral argument, and writing requirements differ from those of lower provincial and federal courts. An emphasis on novel issues of law likely generates new demands for most, if not all, justices. Accordingly, in Table 6 we compare justices' first five years and final five years on the Courts. We replicate the overall variability scores in Table 2, and generate variability scores for the first five and final five years, respectively. The final set of columns report the bootstrap comparison for these two periods. We exclude justices whose total tenure on the Court was fewer than eight years.

A supermajority – 22 out of 29, or 76% - of justices had a higher writing variability during their final five years on the Court. This finding is consistent with the hypothesis that justices, as they reach the latter years of their tenure and the age of retirement, rely more on clerks in the drafting of their opinions. Scholars studying the U.S. Supreme Court have speculated that older justices are prone to “mental decrepitude,”³⁴ given that some justices served well past the age of eighty. Because justices on the Supreme Court of Canada, like other federally appointed judges, must retire by age seventy-five,³⁵ the increased variance in their latter years is likely due, at least in part, to other factors.

33 See generally Lee Epstein et al, *Ideological Drift Among Supreme Court Justices: Who, When, and How Important?*, 101 NW. L. REV. 1383 (2007).

34 See David J. Garrow, *Mental Decrepitude on the U.S. Supreme Court: The Historical Case for a 28th Amendment*, 67 U. CHI. L. REV. 995, 995 (2000).

35 See Constitution Act, 1867, section 99(2).

One explanation for a subset of justices is an increase of administrative responsibilities. For example, four of the five chief justices listed in Table 6 had a higher writing variability in their final five years than their first five years. For three of these four – Chief Justices McLachlin, Lamer, and Dickson – the increase was statistically significant (for Chief Justice Laskin, the increase was more modest and not statistically significant). The exception to this trend was Chief Justice Cartwright, whose variability actually decreased during his final five years, although this difference was not statistically significant. This general trend suggest that the administrative responsibilities that chief justices face may make it more difficult for them to write or draft their own opinions.

TABLE 6
WITHIN-JUSTICE COMPARISON – FIRST FIVE AND LAST FIVE YEARS
CURRENT AND DEPARTED JUSTICES (POST-1948)

Num	Justice	Total			First Five Years			Last Five Years			Bootstrapping		
		Opinions	Avg Word Length	Variability Score	Opinions	Avg Word Length	Variability Score	Opinions	Avg Word Length	Variability Score	P(First5 < Last5)	Lower 95% Confidence Interval	Upper 95% Confidence Interval
5	Morris J. Fish	61	3855	2.98	34	4036	3.03	33	3455	2.84	0.1021	-0.47	0.10
6	Marie Deschamps	61	7283	4.30	25	5931	3.70	36	8222	4.46	0.9997	0.32	1.25
7	Louis LeBel	124	7234	4.58	57	7722	4.57	45	6970	4.50	0.4111	-0.84	0.71
8	Louise Arbour	47	7194	4.13	42	7165	4.17	47	7194	4.13	0.4998	-0.67	0.69
9	William Ian Corneil Binmie	118	7807	3.87	48	8043	3.98	37	6940	3.45	0.0143	-1.01	-0.05
10	Michel Bastarache	84	7794	4.46	38	7749	4.10	36	7812	4.89	0.9983	0.28	1.26
11	John C. Major	101	5113	4.31	29	3876	3.62	41	5445	4.40	0.9982	0.26	1.30
12	Frank Iacobucci	123	8783	5.02	48	8328	4.49	37	8640	5.19	0.9631	-0.07	1.27
14	Beverley McLachlin	304	5529	4.58	91	4144	3.37	53	7109	4.88	1.0000	0.86	2.05
15	Peter Cory	121	6172	4.20	60	5110	3.56	52	7089	4.40	0.9957	0.19	1.40
16	Charles Gauthier	102	7459	4.94	44	8000	4.92	28	7673	4.83	0.3246	-0.82	0.51
17	John Sopinka	145	4476	3.17	73	3616	2.60	72	5348	3.63	1.0000	0.50	1.50
18	Claire L'Heuneux-Dubé	201	6877	4.07	60	7447	4.01	41	6092	3.89	0.3050	-0.74	0.44
19	Gérard La Forest	141	6722	3.77	49	4813	2.79	49	8027	4.11	1.0000	0.79	1.81
21	Bertha Wilson	112	5736	3.97	39	5390	4.09	73	5921	3.84	0.2818	-0.80	0.48
22	Antonio Lamer	228	5541	4.32	36	4529	3.83	38	8306	4.99	0.9985	0.36	1.93
23	Julien Chouinard	56	4234	3.14	36	3885	2.98	28	4900	3.38	0.9755	0.00	0.72
24	William Rogers McIntyre	108	4068	3.16	50	3525	2.92	52	4341	3.11	0.8237	-0.19	0.57
27	Jean Beetz	82	6807	4.47	30	4364	3.70	22	9608	4.69	0.9991	0.34	1.45
28	Robert George Brian Dickson	213	5432	3.72	63	3130	2.89	57	7623	4.24	1.0000	0.82	1.80
29	Bora Laskin	245	3282	3.04	81	2623	2.98	69	3243	3.14	0.6768	-0.39	0.69
30	Louis-Philippe Pigeon	206	2727	2.78	50	2165	2.54	77	3131	2.87	0.9716	-0.01	0.69
31	Wishart Flett Spence	166	3272	3.19	44	2719	2.93	52	3488	2.98	0.6524	-0.27	0.45
32	Emmett Matthew Hall	94	2456	2.98	29	2772	3.01	43	2343	2.92	0.4133	-0.43	0.35
33	Roland Almon Ritchie	286	2856	2.81	51	2544	2.48	28	2698	2.31	0.0208	-0.38	-0.01
34	Wilfred Judson	211	1708	2.62	61	1973	2.62	25	1564	2.48	0.1202	-0.53	0.13
35	Ronald Martland	274	2999	3.81	53	3145	3.75	46	3133	3.25	0.1153	-1.32	0.24
37	Douglas Charles Abbott	121	1183	2.73	27	1095	2.00	27	1180	2.30	0.8768	-0.36	0.76
38	John Robert Cartwright	271	2280	2.82	47	2930	3.12	61	2198	2.45	0.3282	-0.96	-0.35

Note: Table excludes justices who have served fewer than eight years.

Identification of Author: The previous results show that the justices’ writing styles are statistically distinguishable from the null and from each other. These results, in a sense, are an answer to a more modest question than the one asking whether it is possible to use function words to predict which justice authored an opinion. In our informal

conversations with constitutional law scholars, the common response was that they would know the identity of the justice if the passage were well known but as a general matter they could not discern authorship from writing style, particularly of function words.³⁶ To investigate our ability to determine authorship of opinions, we apply a pairwise approach (consistent with our approach for the bootstrap analysis). We compare a universe of opinions written either by Justice A or Justice B.

Intuitively, we determine authorship as follows. First, we compute typical function word frequencies for each justice, based on their other writings. Then, given an opinion of unknown authorship, we see whether its word frequencies are closer to the typical frequencies of Justice A or to those of Justice B. Consider our previous example, in which Justice A uses “this” 2% of the time and “some” 3% of the time, and Justice B uses the word “this” 1% of the time and “some” 4% of the time. In that case, if an opinion of unknown authorship used “this” 2.3% of the time, and “some” 2.9% of the time, then it was probably written by Justice A. By contrast, if an opinion of unknown authorship used “this” 0.8% of the time, and “some” 4.1% of the time, then it was probably written by Justice B. Note that for this author identification, we no longer concern ourselves with frequency variability, but rather with actual word frequencies, which are the most useful way of distinguishing two different writing styles. Of course, this analysis implicitly assumes that each justice has a “typical writing style”, in spite of whatever variability – or assistance from clerks – may be involved. Fortunately, our results indicate that this assumption is reasonable enough, and that we can identify authorship quite successfully using this approach.

To avoid overfitting of the data,³⁷ we divide the data into two parts: training data and testing data. The training data allows us to develop a model for classifying each

³⁶ See Rosenthal and Yoon, *supra* note 16, at 1332-33 (we posed this question of U.S. Constitutional scholars).

³⁷ See Jeffrey S. Rosenthal and Albert H. Yoon, *Detecting Multiple Authorship of United States Supreme Court Legal Decisions Using Function Words*, 5 *Annals Applied Stat.* 283, 287 (describing how overfitting the data is where one constructs a model that fits well on the existing data but cannot effectively predict using new data).

judgment as being written by either Justice A or Justice B. We then use the testing data to see how well our model predicts authorship.

The accuracy of the test is based on *leave-one-out cross-validation*. For each judgment (written by either Justice A or Justice B), this judgment serves as the test data and all other judgments (again written by either Justice A or Justice B) serve as the training data. We identify whether our model accurately attributes the test judgment to either justice. We repeat this process for all judgments written by either Justice A or Justice B, and count the number of accurate classifications.

With a linear classifier, T represents a training set consisting of all judgments by Justice A or B, with $|T| = n$, where n equals the total number of opinions. We use the linear regression model,

$$Y = x\beta + \epsilon,$$

where ϵ is an $n \times 1$ vector of independent errors with mean zero. Y represents an $n \times 1$ vector of : a value of -1 if Justice A actually wrote the opinion, and +1 if Justice B wrote the opinion. The term x is an $n \times 64$ matrix, defined as:

$$x = \begin{pmatrix} 1 & f_{1,1} & f_{1,2} & \cdots & f_{1,63} \\ 1 & f_{2,1} & f_{2,2} & \cdots & f_{2,63} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & f_{n,1} & f_{n,2} & \cdots & f_{n,63} \end{pmatrix},$$

where $f_{i,j}$ are the fraction of words in judgment i in the training set which are from function word j .

The least-squares estimate for β corresponds to the maximum likelihood estimate (MLE), assuming the errors (ϵ_i) to be independent and identically distributed (i.i.d.), defined as

$$\hat{\beta} = (x^T x)^{-1} x^T Y$$

where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n)$. Given a test judgment with function words fractions of g_1, g_2, \dots, g_{63} , the linear fit value corresponds to:

$$\ell = \hat{\beta}_0 + \sum_{j=1}^{63} \hat{\beta}_j g_j$$

If we classify Justice A as the author of the test judgment; if we class authorship to Justice B. We also tried a naïve Bayes classifier, which produces similar, but on average slightly less accurate, predictions to the linear classifier.

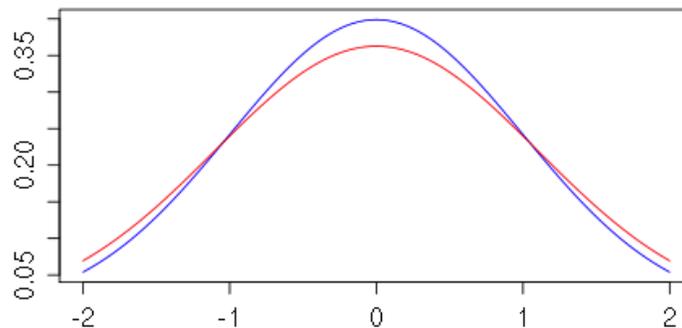
We illustrate how the linear classifier works in Table 7, comparing two pairs of recent justices. An important note regarding interpretation, the pairwise comparison between two justices is not necessarily symmetric. In the comparison between Justice McLachlin with Justice L’Heureux-Dubé, the linear classifier accurately predicted the author of Justice McLachlin 86% of the time, and Justice L’Heureux-Dubé 80% of the time. At the same time, a pairwise comparison can be closely symmetric. The second pairwise comparison between Justice Iacobucci with Justice Sopinka produces a more symmetric result. It accurately predicts Justice Iacobucci as the author 76% of the time and Justice Sopinka as the author 75% of the time. The extent of this symmetry depends on the shape of the probability distributions of the two justices in the pairwise comparison and their degree to which they overlap. If the two distributions largely (but not completely) overlap, it leads to the prediction of authorship to be significantly higher for one of the justices.³⁸

38 The following figure illustrates how the shape and overlap of the distribution between two justices determines the degree of symmetry:

TABLE 7
LINEAR CLASSIFIER
SAMPLE COMPARISONS

Justice A	Justice B	Fraction Accurately Predicting Justice A	Success Rate Predicting Justice A	Fraction Accurately Predicting Justice B	Success Rate Predicting Justice B
Beverly McLachlin	Claire L'Heureux-Dubé	259/302	0.8576	151/188	0.8032
Frank Iacobucci	John Sopinka	93/123	0.7561	107/142	0.7535

Table 8 produces a complete list of pairwise linear classifications from the Lamer Natural Court of 1992-97. Overall, the model does a good job of predicting authorship. Of the seventy-two possible pairings, the model achieved an accuracy rate of at least 70% in fifty-nine pairings (representing 82% of pairings). The lowest pairwise accuracy rate was 0.60, comparing Justice Gonthier (Justice A) with Justice L'Heureux-Dubé (Justice B). Conversely, the highest pairwise comparison accuracy rate was 94%, comparing Justice L'Heureux-Dubé (Justice A) with Justice Cory. As the table illustrates, the prediction rates were higher for some justices than others. For example, the accuracy of prediction for Justice McLaughlin (Justice A) was at least 85% for each of the eight



The blue curve is a graph with mean 0 and variance one. The red curve is a graph with mean zero and variance 1.1. Most of the probability distribution falls between 1 and -1, where the blue curve is larger. In this example, approximately 70% of the points (opinions) chosen from either distribution will be classified as belonging to the blue curve. Accordingly, points from the blue distribution will be correctly classified about 70% of the time, while points from the red distribution will be correctly classified about 30% of the time.

pairwise comparisons; conversely, the accuracy for Justice Major (Justice A) never exceeded 82% and was as low as 55% (pairwise comparison with Justice McLachlin).

TABLE 8
LINEAR CLASSIFIER
SITTING JUSTICES DURING LAMER NATURAL COURT (1992-97)

	Variability Score	Non-Authoring Justice (Justice B)								
		Antonio Lamer	Gérard La Forest	Claire L'Heureux-Dubé	John Sopinka	Charles Gonthier	Peter Cory	Beverley McLachlin	Frank Iacobucci	John C. Major
Antonio Lamer	4.32		0.6950	0.7588	0.8377	0.9035	0.8947	0.8289	0.8333	0.8772
Gérard La Forest	3.77	0.8421		0.6879	0.8085	0.7801	0.8511	0.7589	0.7447	0.8227
Claire L'Heureux-Dubé	4.07	0.7413	0.8209		0.8458	0.8259	0.9353	0.7960	0.7861	0.8756
John Sopinka	3.17	0.6690	0.7724	0.7172		0.8552	0.8138	0.7310	0.7793	0.7862
Charles Gonthier	4.94	0.6569	0.7451	0.5980	0.7549		0.8725	0.5294	0.6373	0.6765
Peter Cory	4.20	0.7521	0.8678	0.8595	0.8347	0.8678		0.7438	0.7851	0.8182
Beverley McLachlin	4.56	0.8454	0.8980	0.8553	0.9079	0.9211	0.9342		0.8849	0.9112
Frank Iacobucci	5.02	0.7154	0.8049	0.6829	0.7561	0.8130	0.8862	0.6341		0.7724
John C. Major	4.33	0.7525	0.7822	0.7624	0.7723	0.7426	0.8218	0.5545	0.6733	

Note: Variability Score column represents score for each justice. Remaining cells represent fraction that accurately predict Justice A compared with Justice B.

It is also important to note that the interpretation from the linear classifier differs slightly from the variability scores. Justices with similar – or even identical – variability scores to one another may nevertheless have clearly distinguishable writing styles from one another. For example, Chief Justice Lamer and Justice Major have nearly identical V scores of 4.32 and 4.33, respectively. The linear classifier, however, predicts with 88% accuracy the authorship of Chief Justice Lamer’s opinions and 75% accuracy the authorship of Justice Major in a pairwise comparison. Similar (or identical) variability scores can nevertheless reflect significant differences in how justices use the various function words. For example, while these two justices use the word *an* at nearly identical rates (mean = 0.61%), Chief Justice Major uses the word *its* (mean = 0.22%) approximately 1.4 times as often as Justice Major.

The purpose of the linear classifier is to show that the text of justices' opinions are often statistically distinguishable from one another, even when the variability scores between two justices do not significantly differ from one another. Of course, identifying the author of a Supreme Court opinion is, in most instances, an academic exercise, given that most majority opinions identify the authoring justice. The linear classifier does, however, have potential applications where the authorship is unknown. *Per curiam* opinions, for example, identify the justices in the majority coalition, but not the author. Similarly, co-authored opinions identify a smaller subset of justices who crafted the opinion, but not the allocation of writing among the justices.

V. DISCUSSION

This article has both a statistical and an institutional objective. The statistical objective was to construct a general approach to evaluating judicial opinions of the Supreme Court of Canada. By relying on common function words rather than case-specific or law-specific terms, we produce measures of the justices' writing variability that allows us to make meaningful comparisons of justices, both within and across individual justices, as well as cohorts. Comparing within justice, over 75% of justices display writing variability that is statistically more variable in their final five years on the Court than in their first five years. And in 83% of pairwise comparisons of the last Lamer natural court, our linear classifier model identifies the correct author at least 70% of the time.

Our second, and arguably more significant objective is to draw inferences from the textual analysis to better understand how justices on the Court produce written opinions. Figure 1³⁹ illustrates that prior to the implementation of judicial clerks, the justices' variability scores were lower, consistent with the view that justices performed more of their own writing. The increased variability – indicated by the upward increased slope in the regression line – is consistent with the view that recent justices rely more on their clerks when writing opinions. It bears noting, however, that the upward slope in

³⁹ See Figure 1, *supra*.

variability scores in Figure 1⁴⁰ over time, prior to the implementation of clerks, suggests an additional factor influencing the use of function words. The most likely explanation is that justices during this period may have increased the practice of quoting other sources – e.g., dicta in other opinions; statutory provisions – that result in increased variability scores.

Also consistent with this point, Figure 2 shows that many justices significantly vary in their variability scores from year to year,⁴¹ reflecting sensitivity in the turnover of clerks. Our variability bootstrap scores provide point estimates of the changes in the variability scores over time.⁴² Chief justices are disproportionately represented among justices who reported a higher variability score the final five years of their tenure than their first five years, which suggests that the administrative demands of being the chief justice makes it more difficult for these justices to focus on opinion writing, meaning that their clerks take on a greater role.

It bears repeating our earlier point that our textual analysis provides only circumstantial evidence of clerks actively participating in the opinion writing stage. Our central premise throughout the article is that the lower a justices' variability score, the less their reliance on their law clerks. We readily acknowledge that even in a world where every justice wrote – without assistance – their own opinions, justices would vary in their variability of writing style. It is also possible that the differences we do observe in variability scores reflect not differences in the behavior of the justices, but in the ability of the clerks to mimic their justices' writing style. These competing explanations, while conceivable, are undermined when looking at the period before the implementation of clerks,⁴³ the year in which Justice Judson wrote opinions without the assistance of a clerk,⁴⁴ and the lower variability of contemporary judges who are widely known to write their opinions without the assistance of their clerks.⁴⁵

40 See Figure 1, *supra*.

41 See Figure 2, *supra*.

42 See Table 5, *supra*.

43 See Figure 1, *supra*.

44 See Figure 3, *supra*.

In our main analysis, we focus on majority opinions rather than concurring or dissenting opinions, for largely technical reasons⁴⁶ due to the fact that concurring and dissenting opinions tend to be shorter than majority opinions. In Figure 9, however, we graphically show variability scores for dissenting and concurring opinions. The close correlation in variability scores, by justice, between concurring and dissenting opinions lends support for the proposition that justices take a more active role when writing these shorter opinions.⁴⁷ They may do so because the advantages of delegation are smaller – i.e., they could write this opinion in the same time that it would take to delegate it to the clerk and subsequently review it – or consumptive value of writing such an opinion is higher.⁴⁸

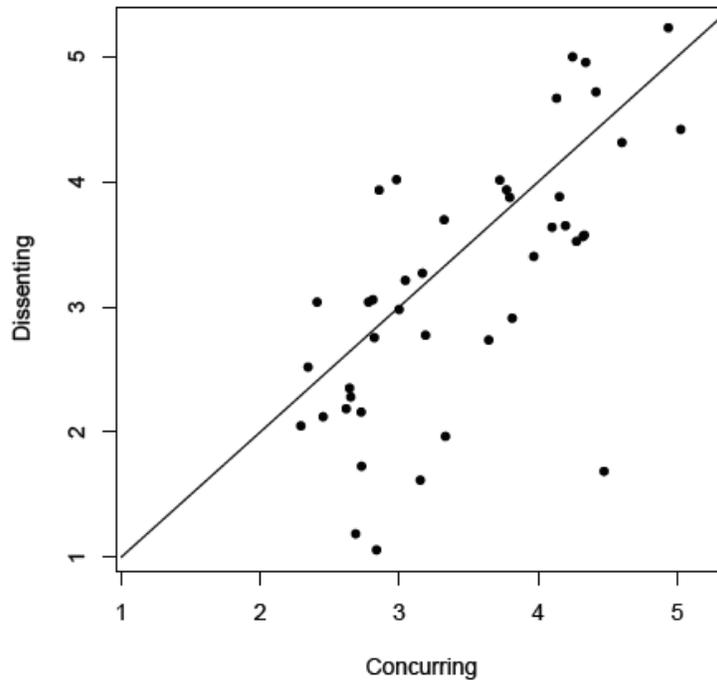
45 See Rosenthal & Yoon, *supra* note 9, at 1325-26.

46 See text accompanying notes 25.

47 This result is consistent with the findings for concurrences and dissents for the U.S. Supreme Court. See Rosenthal & Yoon, *supra* note 36 at 289.

48 See, e.g., “Justice Marshall,” AMERICAN BAR ASS'N JOURNAL, Sept. 1991 at 50 (Justice Marshall, in explaining judicial dissents, commented, “I enjoy the fight. I agree with the old saying, ‘I love peace but I adore a riot.’ You've got to be angry to write a dissent.”).

TABLE 9
VARIABILITY SCORES, CONCURRING AND DISSENTING OPINIONS
BY INDIVIDUAL JUSTICE



Our findings on the linear classifier found that the justices writing style were sufficiently distinct to allow us in the vast majority (82%) of pairwise matching, to correctly identify authorship. It is worth noting that this rate of accuracy, while high, was lower than our analysis of justices on the U.S. Supreme Court, which had an accuracy rate of 94%.⁴⁹ The most plausible explanation for this difference is that the U.S. Supreme Court justices each have *four* clerks, compared with *three* for Supreme Court of Canada justices. The more clerks involved in opinion-writing, all things equal, the greater the variability in writing.

More generally, Supreme Court of Canada justices follow trends similar to the U.S. Supreme Court justices: greater variability in writing with the inception of law clerks, and a general upward trend among current and recent justices. At the same time, the variability scores for U.S. justices are, on the whole, lower than for the Canadian justices. At first blush, this finding seems at odds, given that each U.S. justice has more clerks

⁴⁹ See Rosenthal & Yoon, *supra* note 9 at 1336.

than each Canadian justice. While there are certainly other dimensions in which the two courts are distinguishable – e.g., judicial ideology⁵⁰ - out focus on common function words should be orthogonal to ideology. Canadian justices, while exhibiting a more variable writing style individually, appear more similar to one another along these function words. The explanation, if true, goes beyond this article but merits further consideration.

Much of the criticism targeted towards the U.S. Supreme Court are inapplicable to the Supreme Court. For example, Canadian justices are required to retire at age 75,⁵¹ whereas U.S. justices can serve as long as they like.⁵² Canadian justices may hear cases in panels smaller than nine, allowing them hear more cases if needed; U.S. justices always sit collectively, limiting their ability to manage their docket. The Court, however, does not appear constrained by its docket, as its caseload has been steadily declining in recent years.⁵³

Nevertheless, the common institutional characteristic of both Courts is a bimodal workforce with no middle: senior jurists, typically at least fifty years old, coupled with recent law graduates typically under thirty.

VI. CONCLUSION

This article engages in statistical analysis of judicial opinions in an effort to better understand the Supreme Court of Canada. Our findings provide empirical support to

50 See Benjamin Alarie & Andrew Green, *Should They All Just Get Along: Judicial Ideology: Collegiality, and Appointments to the Supreme Court of Canada* 58 U. NEW BRUNSWICK L. J. 73, 84 (“Given the “brokerage” model of politics in Canada in the past and the lack of significant differences in policy preferences in most areas across parties (particularly in the 1980s and 1990s), the appointees to the Court may have been largely similar ideologically.”).

51 See the Constitution Act, 1867, Part VII (Judicature), Section 99(2).

52 See U.S. Constitution, Article III, Section 1.

53 See Peter McCormick, *Standing Apart: Separate Conurrence and the Modern Supreme Court of Canada, 1984-2006*, 53 MCGILL L.J. 137, 166 (2008) (describing recent caseload trends on the Court).

anecdotal accounts that recent and current justices rely more on law clerks in writing opinions than their predecessors. The normative implications of this finding extend beyond the scope of this article, but it is a discussion worth having, given the import that judges, practitioners, and legal academics attach to the written words of each opinion.

More broadly, this article seeks to build interdisciplinary approaches to understanding jurisprudence and other areas of traditional legal scholarship. The advances in statistical computation and analysis make it possible to systematically explore questions in jurisprudence and doctrine, in a way that can inform how we think both descriptively and normatively about the law.

APPENDIX

TABLE A1 CURRENT AND DEPARTED JUSTICES, PRE-1949

Num	Justice	Province	Appointing Prime Minister	Year Joined Court	Year Left Court	Years On Court	Opinions	Average Word Length	Variability Score
39	Charles Holland Locke	BC	King	1947	1962	15	196	2864	3.32
40	James Wilfred Estey	SK	King	1944	1956	12	214	4331	4.15
41	Roy Lindsay Kellock	ON	King	1944	1958	14	158	2290	2.86
42	Ivan Cleveland Rand	NB	King	1943	1959	16	240	1757	2.45
43	Albert Blellock Hudson	MB	King	1936	1947	11	18	1227	2.49
44	Patrick Kerwin	ON	Bennett	1935	1963	28	293	1650	2.65
45	Henry Hague Davis	ON	Bennett	1935	1944	9	41	2412	2.64
46	Frank Joseph Hughes	ON	Bennett	1933	1935	2	7	2399	2.41
47	Oswald Smith Crockett	NB	Bennett	1932	1943	11	21	2837	3.33
48	Lawrence Arthur Dumoulin Cannon	QC	King	1930	1939	9	7	2003	2.69
49	Robert Smith	ON	King	1927	1933	6	19	1814	2.19
50	John Henderson Lamont	SK	King	1927	1936	9	24	2485	3.36
51	Edmund Leslie Newcombe	NS	King	1924	1931	7	22	2378	2.65
52	Pierre-Basile Mignault	QC	Borden	1918	1929	11	30	1583	2.21
53	Louis-Philippe Brodeur	QC	Laurier	1911	1923	12	32	875	2.05
54	Francis Alexander Anglin	ON	Laurier	1909	1933	24	128	1869	2.41
55	Lyman Poore Duff	BC	Laurier	1906	1944	38	306	4260	3.64
56	Sir Charles Fitzpatrick	QC	Laurier	1906	1918	12	74	1304	2.1
57	James MacLennan	ON	Laurier	1905	1909	4	10	2093	2.5
58	John Idington	ON	Laurier	1905	1927	22	109	1856	2.29
59	Albert Clements Killam	MB	Laurier	1903	1905	2	10	1384	2.03
60	Wallace Nesbitt	ON	Laurier	1903	1905	2	8	2332	2.28
61	John Douglas Armour	ON	Laurier	1902	1903	1	2	574	1.13
62	David Mills	ON	Laurier	1902	1903	1	9	1095	2.05
63	Sir Louis Henry Davies	PE	Laurier	1901	1924	23	108	1772	2.34
64	Désiré Girouard	QC	Bowell	1895	1911	16	17	1310	1.89
65	George Edwin King	NB	Thompson	1893	1901	8	10	1352	2
66	Robert Sedgewick	NS	Thompson	1893	1906	13	25	1998	2.31
67	Christopher Salmon Patterson	ON	Macdonald	1888	1893	5	8	1558	2.23
68	John Wellington Gwynne	ON	Macdonald	1879	1902	23	83	3641	4.34
69	Sir Henri-Elzéar Taschereau	QC	Mackenzie	1878	1906	28	43	1624	2.74
70	William Alexander Henry	NS	Mackenzie	1875	1888	13	46	2634	2.73
71	Jean-Thomas Taschereau	QC	Mackenzie	1875	1878	3	2	2505	5.19
72	Samuel Henry Strong	ON	Mackenzie	1875	1902	27	150	2163	2.73
73	William Johnstone Ritchie	NB	Mackenzie	1875	1892	17	87	1963	3.02
74	Sir William Buell Richards	ON	Mackenzie	1875	1879	4	7	5075	3.25

TABLE A2
VARIABILITY SCORES BY ORIGINAL LANGUAGE OF OPINION

